

PATENT ABSTRACTS OF JAPAN

(43)Date of publication of application : 31.01.2002

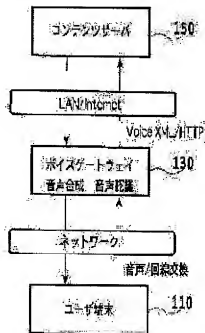
H04M 3/487
G06F 15/00
G10K 15/02
G10L 13/00
G10L 15/00
H04M 3/50
H04M 11/08

(71)Applicant : NTT DOCOMO INC

(72) Inventor: TSUKADA SEIJI

(57)Abstract:

SOLUTION: A user terminal 110 makes voice input and output. A voice gateway 130 recognizes a voice sent from the user terminal 110 and transmits a signal corresponding to the recognition result to a contents server 150. Furthermore, the voice gateway 130 synthesizes voice contents sent from the contents server 150 and transmits the result of voice synthesis to the user terminal 110. The contents server 150 has voice contents and transmits corresponding voice contents to the voice gateway 130 in response to the signal corresponding to the result of voice recognition sent from the voice gateway 130.



(51) Int.Cl. ⁷	識別記号	F I	テラコード ⁷ (参考)
H 0 4 M 3/487		H 0 4 M 3/487	5 B 0 8 5
G 0 6 F 15/00	S 1 0	G 0 6 F 15/00	3 1 0 A 5 D 0 1 5
G 1 0 K 15/02		G 1 0 K 15/02	5 D 0 4 5
G 1 0 L 13/00		H 0 4 M 3/50	A 5 K 0 1 5
15/00		11/08	5 K 1 0 1

審査請求 未請求 請求項の数19 O L (全 14 頁) 最終頁に続く

(21) 出願番号 特願2000-217964(P2000-217964)

(71) 出願人 392026803

(22) 出願日 平成12年7月18日(2000.7.18)

株式会社エヌ・ティ・ティ・ドコモ
東京都千代田区永田町二丁目11番1号(72) 発明者 塚田 晴史
東京都千代田区永田町二丁目11番1号 株
式会社エヌ・ティ・ティ・ドコモ内(74) 代理人 100077481
弁理士 谷 義一 (外2名)

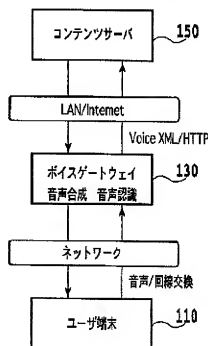
最終頁に続く

(54) 【発明の名称】 サービス提供システム、ボイスゲートウェイ、コンテンツサーバおよびサービス提供方法

(37) 【要約】

【課題】 ネットワークを介したサービスの提供において、ユーザが音声でサービス側とやりとりをできるようにする。

【解決手段】 ユーザ端末110は、音声の入出力を行う。ボイスゲートウェイ130は、ユーザ端末110から送信されてきた音声について音声認識を行い、該音声認識の結果に対応する信号をコンテンツサーバ150に送信する。また、コンテンツサーバ150から送信されてきた音声コンテンツについて音声合成を行い、該音声合成の結果をユーザ端末110に送信する。コンテンツサーバ150は音声コンテンツを有し、ボイスゲートウェイ130から送信されてきた音声認識の結果に対応する信号に応じて、対応する音声コンテンツをボイスゲートウェイ130に送信する。



【特許請求の範囲】

【請求項1】 ネットワーク上でサービスを提供するサービス提供システムであって、
音声の入出力を行うユーザ端末と、
音声コンテンツを有するコンテンツサーバと、
前記ユーザ端末から送信されてきた音声について音声認識を行い、該音声認識の結果に対応する信号を前記コンテンツサーバに送信し、前記コンテンツサーバから送信されてきた音声コンテンツについて音声合成を行い、該音声合成の結果を前記ユーザ端末に送信するボイスゲートウェイとを備え、前記コンテンツサーバは、前記ボイスゲートウェイから送信されてきた音声認識の結果に対応する信号に応じて、対応する音声コンテンツを前記ボイスゲートウェイに送信することを特徴とするサービス提供システム。

【請求項2】 請求項1に記載のサービス提供システムであって、前記ボイスゲートウェイと前記コンテンツサーバとの間で送受信されるコンテンツの変換を行うコンテンツ変換装置をさらに備えたことを特徴とするサービス提供システム。

【請求項3】 請求項1または2に記載のサービス提供システムであって、前記コンテンツサーバは情報コンテンツを有し、前記サービス提供システムは、前記ボイスゲートウェイからの指示に従い、前記ユーザ端末に対して、前記コンテンツサーバから特定の情報コンテンツを受信するように指示するブッシュサーバをさらに備えたことを特徴とするサービス提供システム。

【請求項4】 請求項1または2に記載のサービス提供システムであって、前記コンテンツサーバは情報コンテンツを有し、前記サービス提供システムは、前記ボイスゲートウェイからの指示に従い、前記コンテンツサーバから特定の情報コンテンツを受信し、該情報コンテンツを前記ユーザ端末に送信するブッシュサーバをさらに備えたことを特徴とするサービス提供システム。

【請求項5】 請求項3または4に記載のサービス提供システムであって、前記情報コンテンツは音楽コンテンツであることを特徴とするサービス提供システム。

【請求項6】 請求項1ないし5のいずれかに記載のサービス提供システムであって、
ユーザの認証を行うユーザ認証サーバと、
ユーザに対する課金を管理する課金サーバとをさらに備え、前記ボイスゲートウェイは、前記ユーザ端末が特定の音声コンテンツまたは情報コンテンツを受信する前に、該受信を認めるか否かを判断するために前記ユーザ認証サーバを用いて当該ユーザの認証を行い、前記ユーザ端末が当該音声コンテンツまたは情報コンテンツを受信した後に前記課金サーバに対して当該ユーザに対する課金を指示することを特徴とするサービス提供システム。

【請求項7】 請求項1ないし6のいずれかに記載のサ

ービス提供システムであって、前記コンテンツサーバは、複数のサービス提供言語によるプログラムを有することを特徴とするサービス提供システム。

【請求項8】 請求項1ないし7のいずれかに記載のサービス提供システムであって、前記ユーザ端末は移動機であることを特徴とするサービス提供システム。

【請求項9】 ネットワーク上でサービスを提供するサービス提供システムであって、

音声の入出力を行うユーザ端末と、

10 音声コンテンツを有するコンテンツサーバとを備え、前記ユーザ端末は、入力された音声について音声認識を行い、該音声認識の結果に対応する信号を前記コンテンツサーバに送信し、前記コンテンツサーバから送信されてきた音声コンテンツについて音声合成を行い、該音声合成の結果を出力し、前記コンテンツサーバは、前記ユーザ端末から送信されてきた音声認識の結果に対応する信号に応じて、対応する音声コンテンツを前記ユーザ端末に送信することを特徴とするサービス提供システム。

【請求項10】 請求項9に記載のサービス提供システムであって、前記ユーザ端末は、前記音声認識および前記音声合成を、それぞれ音声認識を行う音声認識サーバおよび音声合成を行う音声合成サーバと通信を行うことにより行うことを特徴とするサービス提供システム。

【請求項11】 請求項9または10に記載のサービス提供システムであって、前記ユーザ端末と前記コンテンツサーバとの間で送受信されるコンテンツの変換を行うコンテンツ変換装置をさらに備えたことを特徴とするサービス提供システム。

【請求項12】 請求項9ないし11のいずれかに記載のサービス提供システムであって、前記コンテンツサーバは情報コンテンツを有し、前記ユーザ端末からの要求に応じて特定の情報コンテンツを前記ユーザ端末に送信することを特徴とするサービス提供システム。

【請求項13】 請求項12に記載のサービス提供システムであって、前記情報コンテンツは音楽コンテンツであることを特徴とするサービス提供システム。

【請求項14】 請求項9ないし13のいずれかに記載のサービス提供システムであって、前記コンテンツサーバは、複数のサービス提供言語によるプログラムを有することを特徴とするサービス提供システム。

【請求項15】 請求項9ないし14のいずれかに記載のサービス提供システムであって、前記ユーザ端末は移動機であることを特徴とするサービス提供システム。

【請求項16】 音声を受信する音声受信手段と、前記音声受信手段により受信した音声について音声認識を行う音声認識手段と、

前記音声認識手段により行った音声認識の結果に対応する信号を送信する対応信号送信手段と、

音声コンテンツを受信する音声コンテンツ受信手段と、前記音声コンテンツ受信手段により受信した音声コン

ンツ音声コンテンツについて音声合成を行う音声合成手段と、
前記音声合成手段により行った音声合成の結果を送信する音声合成結果送信手段とを備えたことを特徴とするボイスゲートウェイ。

【請求項17】 音声コンテンツと、

音声認識の結果に対応する信号を受信する対応信号受信手段と、

前記対応信号受信手段により受信した信号に応答して、
対応する音声コンテンツを送信する音声コンテンツ送信手段とを備えたことを特徴とするコンテンツサーバ。

【請求項18】 ネットワーク上でサービスを提供するサービス提供方法であって、

ユーザ端末において、音声の入力を受ける入力ステップと、

前記入力ステップで入力を受けた音声を、前記ユーザ端末からボイスゲートウェイに送信する音声送信ステップと、

前記ボイスゲートウェイにおいて、前記音声送信ステップで送信した音声について音声認識を行う音声認識ステップと、

前記音声認識ステップで行った音声認識の結果に対応する信号を、前記ボイスゲートウェイからコンテンツサーバに送信する対応信号送信ステップと、

前記対応信号送信ステップで送信した信号に応答して、対応する音声コンテンツを、前記コンテンツサーバから前記ボイスゲートウェイに送信する音声コンテンツ送信ステップと、

前記ボイスゲートウェイにおいて、前記音声コンテンツ送信ステップで送信した音声コンテンツについて音声合成を行う音声合成ステップと、

前記音声合成ステップで行った音声合成の結果を、前記ボイスゲートウェイから前記ユーザ端末に送信する音声合成結果送信ステップと、

前記ユーザ端末において、前記音声合成結果送信ステップで送信した音声合成の結果を出力する出力ステップとを備えることを特徴とするサービス提供方法。

【請求項19】 ネットワーク上でサービスを提供するサービス提供方法であって、

ユーザ端末において、音声の入力を受ける入力ステップと、

前記ユーザ端末において、前記入力ステップで入力を受けた音声について音声認識を行う音声認識ステップと、

前記音声認識ステップで行った音声認識の結果に対応する信号を、前記ユーザ端末からコンテンツサーバに送信する対応信号送信ステップと、

前記対応信号送信ステップで送信した信号に応答して、対応する音声コンテンツを、前記コンテンツサーバから前記ユーザ端末に送信する音声コンテンツ送信ステップと、

前記コンテンツサーバにおいて、前記音声コンテンツ送信ステップで送信した音声コンテンツについて音声合成を行う音声合成手段と、

前記音声合成手段により行った音声合成の結果を送信する音声合成結果送信手段とを備えたことを特徴とするボイスゲートウェイ。

前記ユーザ端末において、前記音声コンテンツ送信ステップで送信した音声コンテンツについて音声合成を行う音声合成ステップと、

前記ユーザ端末において、前記音声合成ステップで行った音声合成の結果を出力する出力ステップとを備えることを特徴とするサービス提供方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ネットワーク上でサービスを提供するサービス提供システムおよび方法等に関する。

【0002】

【従来の技術】従来、インターネット等のネットワークを介してサービスの提供（コンテンツの提供等）が行われていた。

【0003】

【発明が解決しようとする課題】ネットワークを介したサービスの提供において、ユーザが音声でサービス側とやりとりをできれば便利である。例えば、ユーザが音声でサービス側とやりとりをして、音楽コンテンツを購入できれば便利である。多くの携帯電話（移動機）が普及している現在では、そのような音声によるやりとりへの要望も強いと考えられる。

【0004】そこで、本発明の目的は、ネットワークを介したサービスの提供において、ユーザが音声でサービス側とやりとりをできるようにすることである。

【0005】

【課題を解決するための手段】上記目的を達成するために、請求項1に記載の発明は、ネットワーク上でサービスを提供するサービス提供システムであって、音声の入力を行うユーザ端末と、音声コンテンツを有するコンテンツサーバと、前記ユーザ端末から送信されてきた音声について音声認識を行い、該音声認識の結果に対応する信号を前記コンテンツサーバに送信し、前記コンテンツサーバから送信されてきた音声コンテンツについて音声合成を行い、該音声合成の結果を前記ユーザ端末に送信するボイスゲートウェイとを備え、前記コンテンツサーバは、前記ボイスゲートウェイから送信されてきた音声認識の結果に対応する信号に応答して、対応する音声コンテンツを前記ボイスゲートウェイに送信することを特徴とする。

【0006】請求項2に記載の発明は、請求項1に記載のサービス提供システムであって、前記ボイスゲートウェイと前記コンテンツサーバとの間で送受信されるコンテンツの変換を行うコンテンツ変換装置をさらに備えたことを特徴とする。

【0007】請求項3に記載の発明は、請求項1または2に記載のサービス提供システムであって、前記コンテンツサーバは情報コンテンツを有し、前記サービス提供システムは、前記ボイスゲートウェイからの指示に従

い、前記ユーザ端末に対して、前記コンテンツサーバから特定の情報コンテンツを受信するように指示するブッシュサーバをさらに備えたことを特徴とする。

【0008】請求項4に記載の発明は、請求項1または2に記載のサービス提供システムであって、前記コンテンツサーバは情報コンテンツを有し、前記サービス提供システムは、前記ボイスゲートウェイからの指示に従い、前記コンテンツサーバから特定の情報コンテンツを受信し、該情報コンテンツを前記ユーザ端末に送信するブッシュサーバをさらに備えたことを特徴とする。

【0009】請求項5に記載の発明は、請求項3または4に記載のサービス提供システムであって、前記情報コンテンツは音楽コンテンツであることを特徴とする。

【0010】請求項6に記載の発明は、請求項1ないし5のいずれかに記載のサービス提供システムであって、ユーザの認証を行うユーザ認証サーバと、ユーザに対する課金を管理する課金サーバとをさらに備え、前記ボイスゲートウェイは、前記ユーザ端末が特定の音声コンテンツまたは情報コンテンツを受信する前に、該受信を認めるか否かを判断するために前記ユーザ認証サーバを用いて当該ユーザの認証を行い、前記ユーザ端末が当該音声コンテンツまたは情報コンテンツを受信した後に前記課金サーバに対して当該ユーザに対する課金を指示することを特徴とする。

【0011】請求項7に記載の発明は、請求項1ないし6のいずれかに記載のサービス提供システムであって、前記コンテンツサーバは、複数のサービス記述言語によるプログラムを有することを特徴とする。

【0012】請求項8に記載の発明は、請求項1ないし7のいずれかに記載のサービス提供システムであって、前記ユーザ端末は移動機であることを特徴とする。

【0013】請求項9に記載の発明は、ネットワーク上でサービスを提供するサービス提供システムであって、音声の入出力を行うユーザ端末と、音声コンテンツを有するコンテンツサーバとを備え、前記ユーザ端末は、入力された音声について音声認識を行い、該音声認識の結果に対応する信号を前記コンテンツサーバに送信し、前記コンテンツサーバから送信されてきた音声コンテンツについて音声合成を行い、該音声合成の結果を出力し、前記コンテンツサーバは、前記ユーザ端末から送信されてきた音声認識の結果に対応する信号に応じて、対応する音声コンテンツを前記ユーザ端末に送信することを特徴とする。

【0014】請求項10に記載の発明は、請求項9に記載のサービス提供システムであって、前記ユーザ端末は、前記音声認識および前記音声合成を、それぞれ音声認識を行う音声認識サーバおよび音声合成を行う音声合成サーバと通信を行うことにより行うことを特徴とする。

【0015】請求項11に記載の発明は、請求項9また

は10に記載のサービス提供システムであって、前記ユーザ端末と前記コンテンツサーバの間で送受信されるコンテンツの変換を行うコンテンツ変換装置をさらに備えたことを特徴とする。

【0016】請求項12に記載の発明は、請求項9ないし11のいずれかに記載のサービス提供システムであって、前記コンテンツサーバは情報コンテンツを有し、前記ユーザ端末からの要求に応じて特定の情報コンテンツを前記ユーザ端末に送信することを特徴とする。

【0017】請求項13に記載の発明は、請求項12に記載のサービス提供システムであって、前記情報コンテンツは音楽コンテンツであることを特徴とする。

【0018】請求項14に記載の発明は、請求項9ないし13のいずれかに記載のサービス提供システムであって、前記コンテンツサーバは、複数のサービス記述言語によるプログラムを有することを特徴とする。

【0019】請求項15に記載の発明は、請求項9ないし14のいずれかに記載のサービス提供システムであって、前記ユーザ端末は移動機であることを特徴とする。

【0020】請求項16に記載の発明は、ボイスゲートウェイであって、音声を受信する音声受信手段と、前記音声受信手段により受信した音声について音声認識を行う音声認識手段と、前記音声認識手段により行った音声認識の結果に対応する信号を送信する対応信号送信手段と、音声コンテンツを受信する音声コンテンツ受信手段と、前記音声コンテンツ受信手段により受信した音声コンテンツ音声コンテンツについて音声合成を行う音声合成手段と、前記音声合成手段により行った音声合成の結果を送信する音声合成結果送信手段とを備えたことを特徴とする。

【0021】請求項17に記載の発明は、コンテンツサーバであって、音声コンテンツと、音声認識の結果に対応する信号を受信する対応信号受信手段と、前記対応信号受信手段により受信した信号に応じて、対応する音声コンテンツを送信する音声コンテンツ送信手段とを備えたことを特徴とする。

【0022】請求項18に記載の発明は、ネットワーク上でサービスを提供するサービス提供方法であって、ユーザ端末において、音声の入力を受ける入力ステップと、前記入カステップで入力を受けた音声を、前記ユーザ端末からボイスゲートウェイに送信する音声送信ステップと、前記ボイスゲートウェイにおいて、前記音声送信ステップで送信した音声について音声認識を行う音声認識ステップと、前記音声認識ステップで行った音声認識の結果に対応する信号を、前記ボイスゲートウェイからコンテンツサーバに送信する対応信号送信ステップと、前記対応信号送信ステップで送信した信号に応じて、対応する音声コンテンツを、前記コンテンツサーバから前記ボイスゲートウェイに送信する音声コンテンツ送信ステップと、前記ボイスゲートウェイにおいて、前

記音声コンテンツ送信ステップで送信した音声コンテンツについて音声合成を行う音声合成ステップと、前記音声合成ステップで実行した音声合成の結果を、前記ボイスゲートウェイから前記ユーザ端末に送信する音声合成結果送信ステップと、前記ユーザ端末において、前記音声合成結果送信ステップで送信した音声合成の結果を出力する出力ステップとを備えることを特徴とする。

【0023】請求項19に記載の発明は、ネットワーク上でサービスを提供するサービス提供方法であって、ユーザ端末において、音声の入力を受ける入力ステップと、前記ユーザ端末において、前記入力ステップで入力を受けた音声について音声認識を行う音声認識ステップと、前記音声認識ステップで実行した音声認識の結果に対応する信号を、前記ユーザ端末からコンテンツサーバに送信する対応信号送信ステップと、前記対応信号送信ステップで送信した信号に応じて、対応する音声コンテンツを、前記コンテンツサーバから前記ユーザ端末に送信する音声コンテンツ送信ステップと、前記ユーザ端末において、前記音声コンテンツ送信ステップで送信した音声コンテンツについて音声合成を行う音声合成ステップと、前記ユーザ端末において、前記音声合成ステップで実行した音声合成の結果を出力する出力ステップとを備えることを特徴とする。

【0024】以上の構成によれば、ネットワークを介したサービスの提供において、ユーザが音声でサービス側とやりとりをすることができ、

【0025】

【発明の実施の形態】以下、図面を参照しつつ本発明の実施形態について詳しく説明する。

【0026】（第1実施形態）図1は、本発明の第1実施形態に係るサービス提供システムの構成例を示す図である。図1に示すサービス提供システムは、ユーザ端末110、ボイスゲートウェイ130、およびコンテンツサーバ150を備える。

【0027】ユーザ端末110は、音声の入出力を行う。本実施形態において、ユーザ端末110は移動機であるが、固定の電話機を用いることもできる。ユーザ端末110は、例えば、主として音声の送受信を行う従来の移動機であってもよいし、IMT-2000対応の移動機であってもよい。また、入力手段として、キーボード、ポインティングデバイス（マウスなど）等を有するものとすることもできるし、表示手段としてディスプレイ等を有するものとすることもできる。さらに、ブラウザ機能を有するものとする事ができる。

【0028】ボイスゲートウェイ130は、ユーザ端末110から送信されてきた音声（信号）について音声認識を行い、該音声認識の結果に対応する信号をコンテンツサーバ150に送信する（例えば、音声で指定された曲に対応するURL(Uniform Resource Locator)を送信してもよいし、音声認識の結果そのものを送信してもよ

い）。また、コンテンツサーバ150から送信されてきた音声コンテンツについて音声合成を行い、該音声合成の結果をユーザ端末110に送信する。

【0029】コンテンツサーバ150は音声コンテンツを有し、ボイスゲートウェイ130から送信されてきた音声認識の結果に対応する信号に応じて、対応する音声コンテンツをボイスゲートウェイ130に送信する。

【0030】ユーザ端末110とボイスゲートウェイ130の間では、ネットワークを介して音声の送受信が行われる。音声の送受信は回線交換方式により行う。

【0031】ボイスゲートウェイ130とコンテンツサーバ150の間では、ネットワーク（インターネット）を介して（ボイス）サービス記述言語Voice XML(Voice eXtensible Markup Language)に対応したコンテンツの送受信が行われる。コンテンツの送受信はバケット交換方式により行い、プロトコルとしてはHTTP(Hypertext Transfer Protocol)を用いる。

【0032】コンテンツサーバは複数用意することができ、各コンテンツサーバは音声認識機能および音声合成機能を有するボイスゲートウェイ130を介して、ユーザ端末110とやりとりをすることができ、したがって、本システムの構成の方が、各コンテンツサーバに音声認識機能および音声合成機能を持たせる構成よりも経済的である。

【0033】以下、図2～図9を用いて、ユーザ端末110、ボイスゲートウェイ130、およびコンテンツサーバ150の間のやりとりの例を説明する。

【0034】図2は、ユーザ端末110とボイスゲートウェイ130との接続の例を説明するための図である。ユーザ端末110は、ボイスゲートウェイ130と回線交換方式で接続する。

【0035】図3は、サービス選択の例を説明するための図である。ユーザ端末110とボイスゲートウェイ130とが接続されると、ボイスゲートウェイ130はユーザ端末110に、音声「ご利用になるサービス名をご指定下さい」を送信する(1)。これに対してユーザ端末110のユーザが「新譜案内をお願いします」と答え、その音声はボイスゲートウェイ130に送信されたとする(2)。新譜案内が1つ(B1TMUSICの新譜案内)しかない場合、ボイスゲートウェイ130は、音声「B1TMUSICの新譜案内ですね」をユーザ端末110に送信する(3)。このようにして、B1TMUSICの新譜案内が選択され、その新譜案内に関するコンテンツサーバに接続される。新譜案内が2つ以上ある場合には、ユーザにさらに質問して1つを選択させるようにすることができる。

【0036】図4は、曲の選択の例を説明するための図である。コンテンツサーバ150は、新譜案内の要求があると、ボイスゲートウェイ130に音声コンテンツ（その内容は「歌手か曲の名前をご指定下さい」）を送

信する(11)。ボイスゲートウェイ130とコンテンツサーバ150との間のやりとりはパケット交換方式で行われる。ボイスゲートウェイ130は、送られてきた音声コンテンツについて音声合成を行い、その音声合成の結果、すなわち音声(歌手が曲の名前をご指定下さい)をユーザ端末110に送信する(12)。これに対してユーザが「AD」と答え、その音声はボイスゲートウェイ130に送信されたとする(13)。この音声は、ボイスゲートウェイ130で音声認識され、その音声認識の結果に対応する信号がコンテンツサーバ150に送信される。コンテンツサーバ150は歌手および曲(略称等を含む)に関する多くのデータを有しており、「A DJ」が歌手「Abe Defg」の略称であると判断する。ここでは、歌手「Abe Defg」について新譜案内の対象となる曲が「□□□□」のみであるとする。すると、コンテンツサーバ150は、ボイスゲートウェイ130に音声コンテンツ「Abe Defgの「□□□□」ですね」を送信する(14)。ボイスゲートウェイ130は、送られてきた音声コンテンツについて音声合成を行い、その結果をユーザ端末110に送信する(15)。これに対して、ユーザが「はい」と答えたと、その音声はボイスゲートウェイ130に送られ、そこで音声認識されて、その結果に対応する信号がコンテンツサーバ150に送信される(16)。このようにして曲が選択される。

【0037】図5も、曲の選択の例を説明するための図である。図4の(3)のプロセスにおいて、ユーザが「○○○」と答えたとする(図5の(3))。ここで、コンテンツサーバ150に歌手「○○○」について新譜案内の対象となる曲が2曲(「□□□」および「△△△」)あったとする。すると、コンテンツサーバ150は、ボイスゲートウェイ130に音声コンテンツ(「1. 「□□□」と2. 「△△△」の2曲があります。どちらですか。1)を送信する(4))。ボイスゲートウェイ130は、送られてきた音声コンテンツについて音声合成を行い、その結果をユーザ端末110に送信する(5)。これに対して、ユーザが「□□□」と答えたと、その音声はボイスゲートウェイ130に送られ、そこで音声認識されて、その結果に対応する信号がコンテンツサーバ150に送信される(6)。このようにして複数の候補がある場合に1つを選択することができる。

【0038】図6も、曲の選択の例を説明するための図である。複数の候補がある場合には、図4および図5の場合のように音声で選択せずに、ユーザ端末110のボタンで選択するようにしてもよい。例えば、図6に示すように、ユーザは、「1. 「□□□」と2. 「△△△」の2曲があります。どちらですか。」という質問に対して(2)、「1」のボタンを押して「□□□」を選択することができる(3)。

【0039】曲の選択の方法としては、ユーザ端末11

0とボイスゲートウェイ130とをパケット交換で接続し、ボイスゲートウェイ130からユーザ端末110にメニューをブッシュすることも考えられる(図7参照)。曲の選択以外においても、例えばユーザが「ヘルプ」と言ったときにメニューを表示するようにすることができる。

【0040】図8は、試験の例を説明するための図である。曲が選択されると、コンテンツサーバ150は、ボイスゲートウェイ130に音声コンテンツ(「30秒の試験サービスをお聞きになりますか」)を送信する

(11)。ボイスゲートウェイ130は、送られてきた音声コンテンツについて音声合成を行い、その結果をユーザ端末110に送信する(12)。これに対してユーザが「はい」と答えたとする(13)。その音声はボイスゲートウェイ130に送信され、そこで音声認識されて、その結果に対応する信号がコンテンツサーバ150に送信される。コンテンツサーバ150は、ボイスゲートウェイ130に、当該曲の試験ファイルを送信する(14)。ボイスゲートウェイ130は、送られてきた試験ファイル(例えばMP3ファイル)を再生してユーザ端末110に送信する。すなわち試験用音楽を流す(15)。このようにしてユーザは選択した曲を試験することができる。

【0041】図9は、情報提供の例を説明するための図である。コンテンツサーバ150は、ユーザに対して適時に情報提供を行うことができる。例えば、歌手「○○○」の曲が選択された場合に、ボイスゲートウェイ130に音声コンテンツ(「○○○の最新情報をお聞きになりますか」)を送信することができる(11)。ボイスゲートウェイ130は、送られてきた音声コンテンツについて音声合成を行い、その結果をユーザ端末110に送信する(12)。これに対してユーザが「はい」と答えたとする(13)。その音声はボイスゲートウェイ130に送信され、そこで音声認識されて、その結果に対応する信号がコンテンツサーバ150に送信される。コンテンツサーバ150は、ボイスゲートウェイ130に、情報提供の内容、すなわち音声コンテンツ(「大晦日カウント・ダウン5000人ライブ」)を送信する(14)。ボイスゲートウェイ130は、送られてきた音声コンテンツについて音声合成を行い、その結果をユーザ端末110に送信する(15)。このようにしてユーザに対して情報提供を行うことができる。

【0042】図10は、ボイスゲートウェイ130の構成例を示す図である。図10に示すボイスゲートウェイ130は、回線制御部132、音声認識部134、リクエスト作成部136、インターネット接続制御部138、ボイス記述言語パーサ140、音声合成部142、ユーザ情報データベース144、およびローカルデータベース146を備える。

【0043】回線制御部132はユーザ端末110との

間で音声の送受信を行う。インターネット接続制御部138はコンテンツサーバ150との間で、Voice XMLプログラム、音声認識の結果に対応する信号、音声コンテンツ等の送受信を行う。受信したVoice XMLプログラムはボイス記述言語パーサ140で解釈される。

【0044】ユーザ端末110から送られてきた音声は、回線制御部132を介して音声認識部134に入力され、音声認識が行われる。その際、各ユーザの特性等に関するデータを格納したユーザ情報データベース144が用いられる。音声認識の結果は、リクエスト作成部136において、Voice XMLプログラムと比較等されて、対応する信号がコンテンツサーバ150に送信される。

【0045】コンテンツサーバ150から送られてきた音声コンテンツは、インターネット接続制御部138およびボイス記述言語パーサ140を介して音声合成部142に入力される。音声合成部142で音声合成が行われ、その結果を回線制御部132を介してユーザ端末110に送信される。

【0046】ローカルデータベース146には、初期メニューに関するデータ等が格納されている。例えば、図3を用いて説明した新着案内における接続先コンテンツサーバのアドレスが格納されている(図11参照)。

【0047】図12は、コンテンツサーバ150の構成例を示す図である。図11に示すコンテンツサーバ150は、音楽情報データベース(DB)152、音声認識情報データベース154、HTML(HyperText Markup Language)プログラムジェネレータ156、Voice XMLプログラムジェネレータ158、およびWebサーバ160を備える。コンテンツサーバ150は、複数のサービス記述言語によるプログラムを有する。本実施形態においては、HTMLプログラム(コンテンツ)、およびVoice XMLプログラム(コンテンツ)を有する。

【0048】本実施形態におけるコンテンツサーバ150は、HTML用コンテンツサーバの要素(音楽情報データベース152、HTMLプログラムジェネレータ156、およびWebサーバ160)を流用している。すなわち、Voice XMLに対応するため、それらの要素に、音声認識情報データベース154、およびVoice XMLプログラムジェネレータ158を加えた構成をとっている。

【0049】音楽情報データベース152には、歌手(アーティスト)、曲名(タイトル)、曲データ等に関するデータが格納されており、随時更新可能である。例えば、図13に示すようなデータが格納されている。

【0050】HTMLプログラムジェネレータ156は、音楽情報データベース152が更新されると、当該更新が反映されるように現在のHTMLプログラムを

き換える。図14にHTMLプログラムの例を示す。

【0051】音声認識情報データベース154には、読み方、発音等に関するデータが格納されており、随時更新可能である。例えば、図15に示すようなデータが格納されている。なお、「よみ」としては発音記号を用いてもよい。

【0052】HTMLプログラムジェネレータ156は、音楽情報データベース152および/または音声認識情報データベース154が更新されると、当該更新が反映されるように現在のVoice XMLプログラムを書き換える。図16にVoice XMLプログラムの例を示す。

【0053】Webサーバ160は、クライアント(ボイスゲートウェイ130を含む)との間でHTTPプロトコルに基づき、HTMLプログラム、Voice XMLプログラム、各種情報等の送受信を行う。

【0054】(第2実施形態) 図17は、本発明の第2実施形態に係るサービス提供システムの構成例を示す図である。図17に示すサービス提供システムは、ユーザ端末210、ボイスゲートウェイ230、コンテンツサーバ250、およびコンテンツの変換を行うコンテンツ変換ゲートウェイ270を備える。第1実施形態との違いは、ボイスゲートウェイ230とコンテンツサーバ250との間の通信をコンテンツ変換ゲートウェイ270を介して行うことである。ボイスゲートウェイ230とコンテンツ変換ゲートウェイ270との間ではVoice XMLに対応したコンテンツの送受信を行い、コンテンツ変換ゲートウェイ270とコンテンツサーバ250の間ではHTMLに対応したコンテンツの送受信を行う。

【0055】(第3実施形態) 図18は、本発明の第3実施形態に係るサービス提供システムの構成例を示す図である。図18に示すサービス提供システムは、ユーザ端末310、ボイスゲートウェイ330、コンテンツサーバ350、およびブッシュサーバ375を備える。ブッシュサーバ375はボイスゲートウェイ330と一体であってもよい。

【0056】コンテンツサーバ350は情報コンテンツ(本実施形態においては音楽コンテンツであるが、他のコンテンツ、例えば映像コンテンツであってもよい)を有し、ユーザはボイスゲートウェイ330を介したコンテンツサーバ350とのやりとりを通じて情報コンテンツを購入することができる(有料であっても無料であってもよい)。

【0057】情報コンテンツの購入が決まると、ボイスゲートウェイ330はブッシュサーバ375を通じて、ユーザ端末310に、コンテンツサーバ350から当該情報コンテンツを受信するように指示(リクエスト)する。ブッシュサーバ375からユーザ端末310への指示はブッシュプロトコルに基づき行う。ユーザ端末31

13

0はプッシュプロトコルをサポートしている。

【0058】ユーザ端末310は指示を受けると、コンテンツサーバ350と通信を行い、当該情報コンテンツを取得する。コンテンツサーバ350との通信は、HTTPプロトコルを用いてHTMLに基づいて行い、パケット交換方式により行う。

【0059】図19は、コンテンツ購入の例を説明するための図である。例えば、ユーザが曲を選択したとすると、コンテンツサーバ350は、ボイスゲートウェイ330に音声コンテンツ（「購入いたしますか」）を送信する（(1)）。ボイスゲートウェイ330は、送られてきた音声コンテンツについて音声合成を行い、その結果をユーザ端末310に送信する（(2)）。これに対して、ユーザが「はい」と答え、その音声ボイスゲートウェイ330に送信されたとする（(3)）。すると、ボイスゲートウェイ330は、プッシュサーバ375を通じて（図19においてはボイスゲートウェイ330と一体であるとする）、ユーザ端末310に、コンテンツサーバ350から当該情報コンテンツを受信（ダウンロード）するように指示する（(4)）。そして、ユーザ端末310はコンテンツサーバ350と通信を行い、当該情報コンテンツ（音楽ファイル）を取得する（(5)）。このようにしてコンテンツを購入することができ。

【0060】図20は、本実施形態に係るサービス提供システムに、ユーザ認証サーバおよび課金サーバを加えた場合の例を示す図である。ユーザ認証サーバ380は電話番号、ユーザ名等に基づきユーザ認証を行い、課金サーバ385はユーザに対する課金を管理する。ボイスゲートウェイ330は、ユーザ端末310が特定の情報コンテンツを受信する前に、該受信を認めるか否かを判断するためにユーザ認証サーバ380を用いて当該ユーザの認証を行う。また、ユーザ端末310が当該情報コンテンツを受信した後に課金サーバ385に対して当該ユーザに対する課金を指示する。

【0061】図21は、コンテンツ購入の例を説明するための図である。例えば、図19のプロセス(1)の前にユーザ認証を行うことができる（図21の(1)）。また、図19のプロセス(5)の後にユーザに対する課金を行うことができる（図21の(7)）。なお、ユーザに対する課金は、ユーザが情報コンテンツの取得に成功したことを確認した後で行うようにすることが好ましい。

【0062】ユーザ認証は、ユーザ端末310が特定の音声コンテンツ（例えば、情報提供に関するコンテンツ）を受信する前に行うようにしてもよい。また、ユーザに対する課金を、特定の音声コンテンツ（例えば、情報提供に関するコンテンツ）を受信した後に行うようにしてもよい。

【0063】（第4実施形態）図22は、本発明の第4実施形態に係るサービス提供システムの構成例を示す図

14

である。図22に示すサービス提供システムは、ユーザ端末410、ボイスゲートウェイ430、コンテンツサーバ450、およびプッシュサーバ475を備える。プッシュサーバ475はボイスゲートウェイ430と一体であってよい。

【0064】本実施形態においても、コンテンツサーバ450は情報コンテンツを有し、ユーザはボイスゲートウェイ430を介したコンテンツサーバ350とのやりとりを通じて情報コンテンツを購入することができる。

【0065】ただし、本実施形態においては、第3実施形態と異なり、情報コンテンツの購入が決まると、ボイスゲートウェイ430はプッシュサーバ475に、コンテンツサーバ450から当該情報コンテンツを受信し、ユーザ端末410に送信するように指示する。

【0066】プッシュサーバ475とコンテンツサーバ450との通信は、HTTPプロトコルを用いてHTMLに基づいて行い、パケット交換方式により行う。プッシュサーバ475とユーザ端末410との通信は、プッシュプロトコルに基づき行う。ユーザ端末410はプッシュプロトコルをサポートしている。

【0067】なお、本実施形態においても、コンテンツ購入に関連してユーザ認証およびユーザに対する課金を行うことが可能である。

【0068】（第5実施形態）図23は、本発明の第5実施形態に係るサービス提供システムの構成例を示す図である。図23に示すサービス提供システムは、ユーザ端末510、およびコンテンツサーバ550を備える。

【0069】本実施形態においては、第1～第4実施形態と異なり、ユーザ端末510で音声認識および音声合成を行う。

【0070】また、ユーザ端末とコンテンツサーバとの間の通信は、拡張HTML（音声認識タグ付きHTML）またはAV XMLに基づいて行う。拡張HTMLは、HTMLにおいて、リンクに音声認識用タグをつけて、音声認識でリンクをたどれるようにしたサービス記述言語である。AV XMLは、画面表示とボイスガイダンスなど複数の出力手段、音声認識、マウス入力、キーボードなど複数の入力手段を自由に切り替えられるサービス記述言語である。本実施形態においては拡張HTMLまたはAV XMLを用いているが、第1～第4実施形態のようにVoice XMLを用いるようにすることもできる。

【0071】本実施形態においても、コンテンツサーバ550に情報コンテンツを月見し、ユーザ端末510からの要求に応じて情報コンテンツをユーザ端末510に送信するようにすることができる。

【0072】（第6実施形態）図24は、本発明の第6実施形態に係るサービス提供システムの構成例を示す図である。図24に示すサービス提供システムは、ユーザ端末610、コンテンツサーバ650、およびProx

yサーバ670を備える。Proxyサーバ670はコンテンツの変換（音声認識タグの付与等）を行う。第5実施形態との違いは、ユーザ端末610とコンテンツサーバ650との間の通信をproxyサーバ670を介して行うことである。ユーザ端末610とproxyサーバ670との間では拡張HTMLおよびAVXMLに対応したコンテンツの送受信を行い、proxyサーバ670とコンテンツサーバ650の間ではHTMLに対応したコンテンツの送受信を行う。

【0073】（第7実施形態）図25は、本発明の第7実施形態に係るサービス提供システムの構成例を示す図である。図25に示すサービス提供システムは、ユーザ端末710、コンテンツサーバ750、音声認識サーバ790、および音声合成サーバ795を備える。

【0074】第5実施形態では音声認識および音声合成をユーザ端末で行っていたが、本実施形態では音声認識サーバ790および音声合成サーバ795を用いて行う。すなわち、ユーザ端末710は音声認識に必要なデータ（音声データ、認識候補等）をまとめて音声認識サーバ790に送信する。音声認識サーバ790は音声認識を行い、その結果をユーザ端末710に送信する。また、ユーザ端末710は音声合成に必要なデータをまとめて音声合成サーバ795に送信する。音声合成サーバ795は音声合成を行い、その結果をユーザ端末710に送信する。ユーザ端末710と、音声認識サーバ790および音声合成サーバ795との間の通信は、IP(Internet Protocol)プロトコルに基づいて行う。

【0075】なお、ユーザ端末610とコンテンツサーバ650との間の通信は、第6実施形態のようにproxyサーバを介して行うようにしてもよい。

【0076】
【発明の効果】以上説明したように本発明によれば、ネットワークを介したサービスの提供において、ユーザが音声でサービス側とやりとりをすることができる。

【図面の簡単な説明】

【図1】本発明の第1実施形態に係るサービス提供システムの構成例を示す図である。

【図2】ユーザ端末とボイスゲートウェイとの接続の例を説明するための図である。

【図3】サービス選択の例を説明するための図である。

【図4】曲の選択の例を説明するための図である。

【図5】曲の選択の例を説明するための図である。

【図6】曲の選択の例を説明するための図である。

【図7】メニュー表示の例を説明するための図である。

【図8】試聴の例を説明するための図である。

【図9】情報提供の例を説明するための図である。

【図10】ボイスゲートウェイの構成例を示す図である。

【図11】ローカルデータベースに格納されているデータの例を示す図である。

【図12】コンテンツサーバの構成例を示す図である。

【図13】音楽情報データベースに格納されているデータの例を示す図である。

【図14】HTMLプログラムの例を示す図である。

【図15】音声認識情報データベースに格納されているデータの例を示す図である。

【図16】VoiceXMLプログラムの例を示す図である。

【図17】本発明の第2実施形態に係るサービス提供システムの構成例を示す図である。

【図18】本発明の第3実施形態に係るサービス提供システムの構成例を示す図である。

【図19】コンテンツ購入の例を説明するための図である。

【図20】本発明の第3実施形態に係るサービス提供システムに、ユーザ認証サーバおよび課金サーバを加えた場合の例を示す図である。

【図21】コンテンツ購入の例を説明するための図である。

【図22】本発明の第4実施形態に係るサービス提供システムの構成例を示す図である。

【図23】本発明の第5実施形態に係るサービス提供システムの構成例を示す図である。

【図24】本発明の第6実施形態に係るサービス提供システムの構成例を示す図である。

【図25】本発明の第7実施形態に係るサービス提供システムの構成例を示す図である。

【符号の説明】

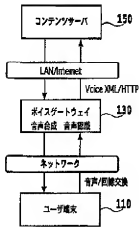
110、210、310、410、510、610、710 ユーザ端末
130、230、330、430 ボイスゲートウェイ
132 回線制御部
134 音声認識部
136 リクエスト作成部
138 インターネット接続制御部
140 ボイス記述言語サーバ
142 音声合成部
144 ユーザ情報データベース
146 ローカルデータベース
150、250、350、450、550、650、750 コンテンツサーバ
152 音楽情報データベース
154 音声認識情報データベース
156 HTMLプログラムジェネレータ
158 VoiceXMLプログラムジェネレータ
160 Webサーバ
270 コンテンツ変換ゲートウェイ
375、475 プッシュサーバ
380 ユーザ認証サーバ
385 課金サーバ

670 proxyサーバ
790 音声認識サーバ

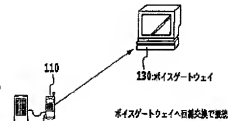
* 795 音声合成サーバ

*

【図1】



【図2】



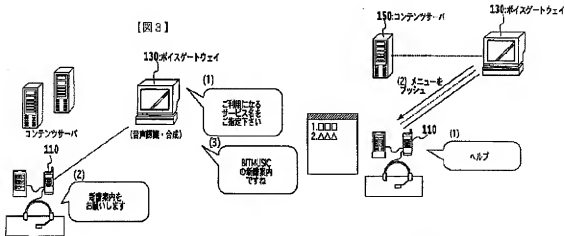
ボイスゲートウェイへ目覚め交換で接続

—— 目覚め交換
—— パケット交換

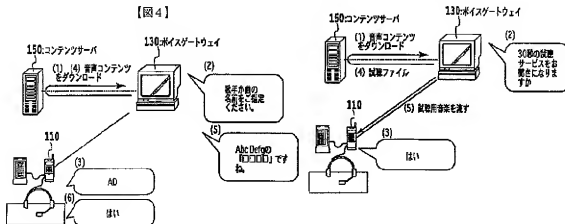
【図3】

アーティスト	タイトル
AOC	想いみに咲く花
エスラフ/ックス	美女
岡田晴幸	LIFE
辻藤和彦	恋はじかじか
早部	Motto

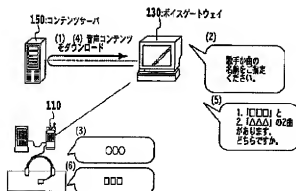
【図7】



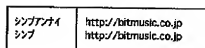
【図8】



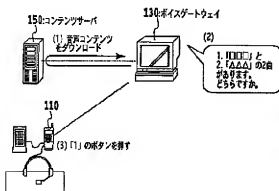
【図5】



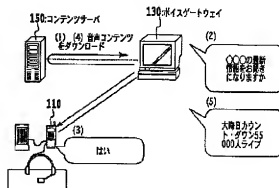
【図11】



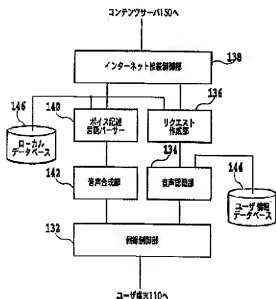
【図6】



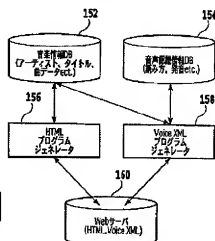
【図9】



【図10】



【図12】



【図14】

```

<?xml>
<HEAD><TITLE>MUSIC NAVI</TITLE></HEAD>
<BODY>
◆新機能情報
<FORM ACTION="/cgi-bin/mode/search/search.cgi"
NAME="search" METHOD="POST">
<SELECT NAME="mode">
<OPTION VALUE="artist">アーティスト
<OPTION VALUE="title">タイトル
</SELECT>
<BR>
<A HREF="/single/9912/0001.htm">AOC 魅しみに度く
花</A>
<A HREF="/single/9912/0001.html">エスラフノックス
"天使"</A>
<A HREF="/single/9912/0001.html">岡田唯幸 "UES"
</A>

```

HTMLプログラム

【図16】

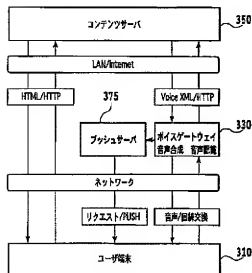
```

<?xml version="1.0">
<xsl:stylesheet>
<form>
<field name="artistarticle">
<prompt count="1">最中のお曲名を指定してください。</prompt>
<prompt count="2">AOC エスラフノックス、岡田唯幸、佐藤和真、平井が登録されています。</prompt>
<grammar>
<rule>
<alt>
<alt>
<alt>
<alt>
<alt>
<alt>
<alt>
</alt>
</rule>
</grammar>
</form>
</xsl:stylesheet>

```

Voice XMLプログラム

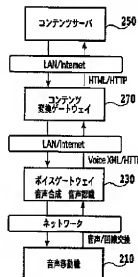
【図18】



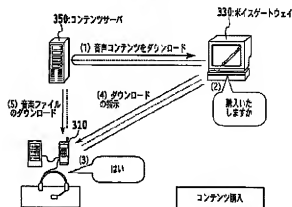
【図15】

アーティスト	名前
AOC	スーパースター
エスラフノックス	えすらのっくす
岡田唯幸	おかだやすき
佐藤和真	さとうかずま
平井	ひら

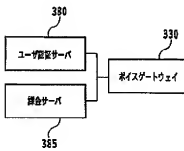
【図17】



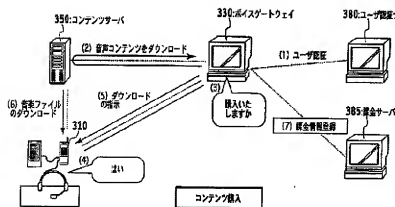
【図19】



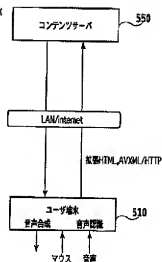
【図20】



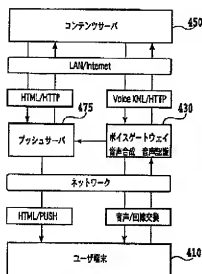
【図21】



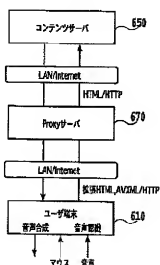
【図23】



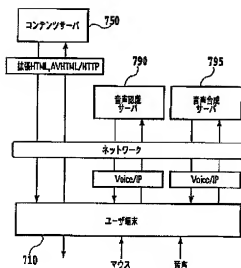
【図22】



【図24】



【図25】



フロントページの続き

(51)Int. Cl. ⁷	識別記号	F I	テーマコード(参考)
H 0 4 M	3/50	G 1 0 L	3/00 R
11/08			5 5 1 A

Fターム(参考) 5B085 AA08 AE23
 5D015 KK02
 5D045 AB01
 5K015 AA06 AA07 AD01 AD02 AD05
 GA07 GA09
 5K101 KK16 KK17 LL12 MM04 MM07
 NN07 NN08 NN16 NN22 NN25
 TT03